

Universitat de Lleida

Document downloaded from:

<http://hdl.handle.net/10459.1/71243>

The final publication is available at:

<https://doi.org/10.1007/s10209-019-00660-3>

Copyright

(c) Springer-Verlag, 2019

Mireia Ribera*, Ricardo Pozzobon, Sergio Sayago

Publishing accessible proceedings: the DSAI 2016 case study

Universitat de Barcelona. Departament de Matemàtiques i Informàtica. Gran Via de les Corts Catalanes 585, 08007 Barcelona

Universitat de Barcelona. Departament de Biblioteconomia i Documentació. Estudiant. Melcior de Palau, 140, 08014 Barcelona

ribera@ub.edu, 934020855 orcid.org/0000-0003-1455-1869

Ricardo Pozzobon orcid.org/0000-0001-7043-9711

Sergio Sayago orcid.org/0000-0002-5227-1246

Acknowledgments

Thanks to Hugo Paredes, during and after the conference, for his support and help, and for carefully reviewing this article.

This is a post-peer-review, pre-copyedit version of an article published in Universal Access Information Society. The final authenticated version is available online at: <https://doi.org/10.1007/s10209-019-00660-3>

Publishing accessible proceedings: the DSAI 2016 case study

Mireia Ribera*, Ricardo Pozzobon, Sergio Sayago

Received: date / Accepted: date

Abstract Access to research papers has changed in the last decades: from printed to digital sources, from close to open access. Despite these changes and broader access to research results, there are still accessibility barriers. A very few number of conferences and journals state an accessibility policy of their publications in their websites. The production of accessible documents is not a common practice in conferences devoted to accessibility. Purpose: The purpose of this paper is to present the case study of DSAI 2016 (Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion) conference, wherein we were in charge of making accessible its proceedings. Methods: We discuss the methods and technical procedures we carried out to turn the original articles in MS Word and Latex formats into accessible PDFs and the steps necessary in authoring, conversion and validation. Results: the papers of DSAI 2016 were published in accessible format after much effort, best tools and procedure were using MS Word plus PDF Axes plus PDF Accessibility checker. Conclusion: We state the need to include a new role in the organizing committee of conferences for dealing with accessible publishing.

Keywords Document accessibility, Accessible proceedings, PDF/UA, EPUB3, Conversion procedures

1. Introduction

Academic conferences should be accessible by default [1] [2]. Otherwise, the dissemination of scientific knowledge for all is seriously hindered. In this paper, we concentrate on the technical accessibility of conference proceedings. By accessibility we do not mean free access to conference proceedings. Accessibility, in this paper, means being able to read the papers published in the conference proceedings regardless of any disability or impairment you might have, or software you use. In addition to being the main scientific output of a conference, the proceedings are often part of the bibliography for higher university courses, and a basic source of information for researchers. Thus, publishing in accessible formats is of the utmost importance, and a matter of respect, compromise and coherence, especially for those conferences dealing with accessibility and digital technologies. Several conferences have promoted initiatives to foster the publication of accessible proceedings. TEX Users group (TUG) has also been active working on this issue. However, the results still are not consolidated, and articles are not accessible [3].

In this paper, we reflect on our own experience of publishing the proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2016) conference in an accessibility format. We share what we learned about the possibilities and limitations of authoring tools, such as MS Word, InDesign, Acrobat DC and PDF Axes for Word, to create accessible proceedings or convert them into an accessible format. We argue that publishing accessible proceedings is not straightforward and that it requires a lot of planning at the submission stage. We highlight the importance of an effective communication between the authors and the publishing team to speed up the generation of accessible proceedings. We also indicate that Latex sources still are not correctly solved with the current tools and technologies.

Although it is difficult to make general claims from case studies, the lessons we learned encourage us to argue that generating accessible conference proceedings in an easy, efficient and effective way is currently a very difficult objective to achieve, and that we need an integrated workflow to make it a more doable task. Our suggested workflow consists of four key elements: better tools, better communication between conference organization and researchers, better templates and instructions, and more awareness of accessibility.

The paper is organized as follows. In the Background section, we provide an overview of tools, formats and technical procedures. Afterwards, in the Related Research section, we review and discuss previous initiatives related to this paper. We then present the case study reflect on the lessons learned and the actions taken, and show our proposal. We end up with a discussion, conclusion and future research perspectives.

2. Background

This section aims to provide a background on accessible publishing. In 2.1, we give a succinct overview of the main digital formats and tools used to author accessible documents. Afterwards, we summarize three key steps in the publishing process: preparation of originals, conversion and validation. Table 2 presents a summary of steps and tools.

2.1 Format and tools

The technical format of a digital document establishes a threshold on its accessibility, which is determined by authoring and reading tools [16]. For example, HTML is a very flexible format, because HTML files separate the content from the presentation. The complete source of the file can be seen while authoring it, which gives the authors high control over the final results. In what follows, we review the most common formats in which conference proceedings are usually published: PDF, TeX, and MSWord. We also turn our attention to a relatively new file format: EPUB3.

2.1.1 PDF

The main goal of the Portable Document Format (PDF) is to offer presentation reliability through different platforms and printing devices. The company's strategy to offer Adobe Reader freely and to actively listen to the prepress professionals' needs did pay back with a big adoption against competing formats such as Dejavu. PDF was a proprietary format until 2008, when led by the pressure of governments to use open formats, it became an ISO standard. A further step happened in 2012 when, again caused by law requirements' pressures, it incorporated accessibility in its variant PDF/UA [17]. PDF/UA¹ is a new standard coherent with WCAG 2.0 guidelines, and it is detailed in ISO 14289-1:2012 (version 1) and ISO 14289-2 (version 2) standards [12]. Following the changes within the format, Adobe Reader has incorporated some worth mentioning accessibility functions, such as the possibility to resize text, reflow the content as a unique column, change colours, automatically play forward or back the pages, and, if the platform provides a screen-reader API feature, to read-aloud the content.

The main limitation of PDF lies in its binary format, which is not suitable for a raw authoring procedure such as Latex or HTML. There are tools, though, such as Framemaker, which work with XML to create PDF documents. On 2009, Adobe labs released the Mars Project to offer an XML version of the format. However, the project was discontinued and no further steps have been taken at the time of writing this paper. Due to the PDF origins, many authoring tools only focus on the layout and presentation, and do not incorporate either the new semantic structure created since PDF version 1.7 or the new elements introduced in PDF/UA. Despite these shortcomings, PDF is widely regarded as the most reliable publishing format in every platform. Adobe Reader is a reading tool offering flexibility for creative layouts, different fonts and all kinds of special symbols, with a perfect presentation wherever it is used, except for mobile versions of Adobe Reader which, unfortunately, do not offer reflow capabilities. Adobe released in 2017 a new version of its main standard, Document management – Portable Document format Part 2: PDF 2.0 [17], which is still in development.

2.1.2 Docx

¹ <http://www.digitalpreservation.gov/formats/fdd/fdd000350.shtml>

The Open office document format is generated by MS Word, from the Microsoft Office suite. Docx is one of the most widely used formats. Since 2010, MS Word includes an accessibility wizard that automatically checks, among other features, if there are alternative texts and if the headings hierarchy is correct. Docx is rarely used as an output format, because the appearance may change depending on the platform or device used. Therefore, it is common to create a PDF document as the final version of a MS Word document. By default, MS Word offers a conversion tool to PDF. Yet, this conversion is not perfect and the final document does not follow PDF specifications. Another way to generate a PDF from a docx document is to add a plugin from Adobe Acrobat in MS Word. Although this plugin helps authors to improve the quality of the resultant PDF, this procedure will not either create a compliant PDF/UA document. As mentioned before, there are other tools which are especially geared towards converting from docx to PDF/UA documents, mainly Axes PDF² and NetCentric CommonLook³.

2.1.3 TEX

TEX is a typesetting system designed by Donald Knuth in 1978 in response to two needs: the production of high-quality books and the portability of documents, both with free software. TEX is mainly used to create documents with mathematical formulas. TEX consists in plain text with marks to indicate semantics. The output of a TEX document is usually a PDF document. LaTeX is a widely used document writing system. LaTeX aims to simplify TEX document authoring and is free. Since its creation, TEX has not undergone big changes, because backward compatibility is more valued than evolution. This lack of evolution is both its main advantage and disadvantage. This implies, for example, that while it is possible to read a 30 year-old document today, it is impossible to include any videos, audios or activities, such as quizzes. TEX users, who are spread worldwide, have developed and made available numerous libraries to implement specific functions within TEX. Such a large number of libraries might create incompatibility problems. As an example, to avoid incompatibility issues, a specific TEX library to include accessibility requirements was used in CHI 2014 (Computer Human Interaction conference).

2.1.4 Electronic Publication Format (EPUB)

EPUB was created in 1999 by the International Digital Publishing Forum (IDPF) – named Open eBook forum at that time – to avoid the standard fragmentation in ebook formats. This fragmentation forced publishers to publish their contents in several proprietary formats. In Switzerland, in parallel, a group of researchers worked together to define a format suitable for very different needs, in terms of both presentation and interaction. This work led to the Digital Accessibility Information System (DAISY) format (Petrie, Weber, 2002). DAISY aimed to be a format suitable for any type of document. DAISY gained momentum in 2000 (Ribera & Moese 2008) but it was not widely adopted, as it was very specific for blind or low vision people. In 2011, IDPF and DAISY Consortium joined forces and created a new version of the EPUB technical specification, EPUB3. The new version of EPUB brought together the best of both worlds: a rich and flexible format for publishers plus a format adaptable to many different user profiles. The goals were to get maximum adoption and to become the preferred format for small and big publishers, looking towards mainstreaming accessibility.

EPUB3 versatility is making it a suitable format for teaching and learning. In 2011, EPUB 3 gained the endorsement of the Publishing Industry as the single standard to rely on. Additionally, EPUB3 is the format of choice by the accessible books consortium [19], which is led by the World Intellectual Property Organization (WIPO), including World Blind Union, International Authors Forum, IFLA and International Publishers Association. IDPF's recent merge with W3C is another sign of this trend to become a full standard for the digital world. Yet, there is a lack of authoring and conversion tools. For example, MS Word does not offer a process to convert docx or doc documents to EPUB. To do so, we have to Save as Daisy, by adding the Add-In Word plugin⁴ by the DAISY Consortium. This option generates a DAISY Digital Talking Book (DTB) [14] document, which can be converted to EPUB3 using Tobi⁵ desktop tool [20].

2.1.5 Images

² <https://www.axes4.com/axespdf-for-word-overview.html>

³ <http://commonlook.com/accessibility-software/commonlook-office-globalaccess/>

⁴ <http://www.daisy.org/project/save-as-daisy-ms-word-add-in>

⁵ <http://www.daisy.org/project/tobi>

Docx format do not cover vector images, which offer many advantages for zooming and are especially suitable for low vision users. Docx relies mainly on bitmap images. Bitmap most standardized formats are GIF, PNG and JPEG. JPEG is the most similar to PDF internal image format and does not include a transparency layer, avoiding some occlusion problems when customizing colours in Adobe Reader.

2.2 Publishing process

Three key steps are involved in publishing accessible conference proceedings, namely preparation of originals, conversion to the publishing format and validation. The second and third steps can be done by either authors or the organizing committee or an outsourced service.

2.2.1 Preparation of originals

Manuscripts tend to be authored in either MS Word or LaTeX, and then converted to PDF. As mentioned above, the PDF generated by LaTeX is not accessible and must be repaired. There are better tools for the conversion from MS Word to PDF, such as Axes PDF and NetCentric CommonLook.

Most accessibility flaws are often due to the template for documents provided to the authors. Examples are tables and images without alternative text, an incorrect hierarchy of titles, and a lack of embedded fonts. The template, which usually gives instructions on how to write papers, does not usually come with accessibility recommendations. Thus, the original documents do not have metadata, use tables as layout tools, and tend to miss alternative texts for images. All this information is essential for generating accessible documents.

For LaTeX documents, one of the possible solutions to create accessible PDF documents is to use the Infty Editor, a scientific optical text recognition that creates accessible PDF from an image or from a PDF file. We used this editor with real authored articles. The authors found that this tool identifies many textual contents as formulas, making it difficult to process the resulting PDF documents. Another possible solution is to repair the converted PDF with Adobe Acrobat editing capabilities. However, this solution is tedious for complex documents, because the interface for editing tags is not usable. Finally, TEX documents can be repurposed as MS Word documents; in this case you may find that author templates for MS Word and LaTeX are not the same. Thus, the repurposing is not straightforward; formulas and vector images require specific conversions too, and the conversion may imply some layout changes.

For MS Word documents, if the tool settings are not configured properly, images are downsized to low resolution. Missing information such as alternatives, fonts or metadata, are common problems in generating accessible documents from doc or docx originals.

2.2.2 Conversion

The conversion from MS Word to PDF is not straightforward, even with professional tools such as Adobe Acrobat. Adobe® Acrobat DC is a full editing tool that allows us to create a compliant PDF/UA document. Adobe® Acrobat DC offers complete functionalities to edit and delete the semantic structure of a document. However, the conversion plugin for MS Word does not follow PDF/UA requirements and to comply with them you should rely on a post-hoc manual editing, which is a tedious process. Authors trying to create PDF/UA documents are overwhelmed with complex and time-consuming repairing tasks on the final documents [13]. For example, the tags created by default in the conversion procedure (see table 1) are a reduced set of the more heterogeneous possibilities PDF/UA offers to the author.

Table 1. PDF Tags related to proceedings content and default generated tag by Adobe Acrobat MS Word plugin

Type of content	Required Tag	Default generated tag
Article	<Art>	--
Annotation	<Annot>	--
Bibliography	<BibEntry>	<L> ...
Blocked Quote	<BlockQuote>	-
Caption	<Caption>	<Caption>
Code	<Code>	-

Division	<Div>	-
Document	<Document>	-
Figure	<Figure>	<Figure>
Formula	<Formula>	<Formula>
Heading	<H>	<H>
Heading level 1-6	<H1>...<H6>	<H1>...<H6>
Index	<Index>	-
Link	<Link>	<Link>
List	<L>,,<Lbl><LBody>	<L>,,<Lbl><LBody>
Note	<Note>	-
Paragraph	<P>	<P>
Part	<Part>	<Part>
Quote	<Quote>	-
Reference	<Reference>	-
Section	<Sect>	-
Span		
Table	<Table><TR><TD><TH>	<Table><TR><TD><TH>
Artifact	<Artifact>	<Artifact>

A very important failure in the conversion process is the missing <document> tag, which is the root element of any PDF document structure.

Another possibility to create PDF documents from MS Word is to use Adobe® InDesign. *Adobe® InDesign* is the main authoring tool for PDF documents. It offers functionalities to design the layout and style of any document and to include content from other applications, in particular from MS Word. Adobe® InDesign offers the possibility to create PDF and EPUB documents by default. The procedure to create an InDesign document from a MS Word document consists in pasting separately text, images and tables. If the layout comprises columns, these should be pasted separately on the final InDesign document or recreated once all the content is pasted. If the original MS Word documents has the content correctly marked with styles, these are imported within Adobe InDesign and can be automatically converted to structural tags. A good way to do this is to create a Word template (see for example a tentative proposal for ACM conferences in ⁶) and also an InDesign template (see the ACM InDesign counterpart in ⁷) with structural tags that have the same names so that the styles will automatically be converted to the corresponding tag. The template can also be populated with a prerecorded list of tags, such as the one suggested in ⁸. There are some settings that must be customized in order to create a PDF/UA document, these can be defined in a joboptions file or in the metadata, where you have to set author and title (this information is not retrieved from the original Word document) and also the PDF/UA rdf description⁹.

The main drawback of the conversion from InDesign to PDF is, again, the reduced number of tags generated. Although the tags keep their names, all the tags except for lists, tables and links, which are automatically recognized and tagged correctly, are repurposed to P, H, H1, H2, H3, H4, H5, H6 internally. If you want to use the original or a more complete tag set, you must post-process the documents (either one by one or by a script) and map the names to the correct tag.

AxesPDF is another tool to convert documents from MS Word to PDF, especially designed with accessibility in mind. The converted files follow the PDF/UA-1 standard. AxesPDF is a tool that takes the author in consideration and streamlines some of the cumbersome steps of the conversion process by, for example, embedding the fonts, including PDF/UA xml metadata, and converting table borders to artifacts. The conversion process can be customized within the tool to establish particular conversion preferences [20]. Although this tool does not cover many tags, apart from the basic ones, it includes Caption, and BlockQuote, which are handy for conferences, because conference proceedings are full of images, and often include scientific quotes.

⁶ MS Word ACM tentative template from <http://www.ub.edu/adaptabit/acm/pubform-accessible.dotx>

⁷ InDesign ACM tentative template from <http://www.ub.edu/adaptabit/acm/pubform-accessible.indt>

⁸ Definition of tags for a paper <http://www.ub.edu/adaptabit/acm/tags.xml>

⁹ XMP metadata file required for PDF/UA <http://www.ub.edu/adaptabit/acm/pdfUA.xmp>

2.2.3 Validation

Validation is a mandatory step when you want to obtain compliant PDF documents. Authoring tools may not follow the required specifications. Validation could be done at any stage of the authoring process. However, given that detecting an error later on in the process might entail doing again a large number of steps, it is common to validate the generated documents as soon as possible. For example, it is a good practice to do the first validation in the original Word documents. This can be done by using the MS Word accessibility wizard. This first validation can help authors detect, for example, failures in providing alternative texts, in tables and in headings hierarchy.

After the conversion to PDF, a second validation step is necessary. This step consists of following the Matterhorn Protocol¹⁰, which includes 31 checkpoints and 136 fail conditions. From these 136 conditions, 108 can be automatically validated, and the others require a manual review. The automatic review can be done by Adobe Acrobat Accessibility validation tool or by specialized tools such as PDF Accessibility Checker. PDF Accessibility Checker version 2.0 (PAC2) is free, promoted by “Access for All” Swiss foundation¹¹, and developed by xyMedia¹². PDF Accessibility Checker allows us to technically validate PDF documents against the PDF/UA Standard. The tool gives aggregated and detailed results of the checkpoints, and offers a detailed view of the document section highlighting errors. It also offers the possibility to view the document with structural tags embedded to evaluate the correct semantics and reading order.

Preparation	Conversion	Conversion comments	Validation	Validation comments
Word	Acrobat Professional	Does not create PDF/UA documents by default. Posterior repairing requires time.		
Word	InDesign	Requires some additional work to create the original, generates a reduced tag set. Converts to PDF and to EPUB.		
Word	AxesPDF	Easy process, more flexible. Reduced tag set.		
Latex	Infty Editor	Creates false formulas.		
Word			Word Accessibility Wizard	Identifies errors at an early stage
PDF			Adobe Acrobat	Validation and repairing. Does not cover many PDF/UA requirements during validation
PDF			PDF Accessibility Checker	The most complete tool. Only to validate

Table 2. Summary of tools for preparing, converting and validating accessible conference proceedings

3. Related research

¹⁰ <https://www.pdfa.org/publication/the-matterhorn-protocol-1-02/>

¹¹ <http://www.access-for-all.ch/>

¹² <http://www.xymedia.ch/>, xy Media is a member of PDF competence center

As stated in [3], the accessibility level of conference proceedings, even those related to accessibility, is not satisfactory and does not comply current standards. There are numerous barriers, such as access to formulas or missing alternative texts which hinder access to conference proceedings by blind people, people with dyslexia, and people who rely on computers to read aloud or interpret content.

ACM SIGCHI was the first to promote accessibility requirements in their conferences. In CHI 2014, the sigchi-latex Google Code Project¹³ was an initiative to create accessible PDFs from Latex sources. This initiative was possible thanks to Schalit, who created an “accessibility” Latex package for covering alternate texts for images and table headers. This package allows us to generate tagged PDFs, which indicate the default language and the reading order. Clifton updated this package and fixed some bugs. Unfortunately, this project has remained inactive since 2012. Within the sigchi-latex code project, the SIGCHI 2014 organizing committee created specific MS Word and Latex templates with additional information for accessibility, and these have been updated regularly on GitHub¹⁴, with the last commit on January 2017.

After this SIGCHI initiative, the Tex users group (TUG) has devoted a specific page to show the resources needed for generating accessible PDFs [4]. TUG has also recently opened an active discussion on the conversion from TEX to PDF. This conversion is not either reliable or comprehensive yet¹⁵. Mainstream LaTeX authoring tools, such as MikTeX, generate by default a PDF without semantic structure. Without this structure, a PDF is not technically accessible. Moreover, the generated PDFs tend to miss blank spaces between words and have many reading order problems. Current research focuses on experimental tools¹⁶ (refer to TUG mailing list for more information). Another option is to convert LaTeX documents into HTML or EPUB. This option has received a lot of interest (5-10%) but currently does not solve conference proceedings.

In CHI 2015, a group of researchers volunteered to convert camera-ready PDF versions to accessible versions. They created a guide for authors. This guide asked the authors to provide them with alternate descriptions and embedded fonts, amongst other aspects. The authors had to generate accessible PDFs of their own papers and send them to the volunteers, who reviewed and repaired them. This experience is described in [3], where further information about time and costs of repairing and possible ways to outsource this work are discussed.

In 2016, some of the authors of [3], after additional experiences of creating and repairing accessible PDF proceedings, reflected on the complexity of the PDF format. In their view, it is not easy for neither authors nor the conference committee to create or afford to pay for accessible papers [11].

On a minor level of compromise, other conferences have published recommendations for authors, like ASSETS¹⁷ or ICCHP¹⁸, with basic instructions on how to generate better PDFs from MS Word. SIGACCESS¹⁹ provides a style guide on how to correctly mention in scientific publications disabilities and people with disabilities. CHI 2019²⁰ has recently published a detailed guide for accessible submissions, but Aptara template and authors’ workflow proposal has been a failure and has recently (July 2018) been pulled.

It is worth mentioning that some journals in HCI, such as Universal Access to Information Society²¹, have also included specific accessibility recommendations in the instructions for authors, while others, such as RESNA²², have not done so yet.

¹³ <https://code.google.com/p/sigchi-latex/wiki/Accessibilityconferences>

¹⁴ <https://github.com/sigchi/Document-Formats>

¹⁵ <http://tug.org/pipermail/accessibility/2016q4/thread.html>

¹⁶ <http://tug.org/pipermail/accessibility/2018q3/000048.html>;

<http://tug.org/pipermail/accessibility/2018q3/000046.html>

¹⁷ <http://www.sigaccess.org/welcome-to-sigaccess/resources/accessible-pdf-author-guide/>

¹⁸ <http://www.icchp.org/content/impact-pdfua-accessible-pdf>

¹⁹ <http://www.sigaccess.org/welcome-to-sigaccess/resources/accessible-writing-guide/>

²⁰ <http://chi2019.acm.org/authors/papers/guide-to-an-accessible-submission>

²¹ <http://www.springer.com/computer/hci/journal/10209>

²² <http://www.tandfonline.com/toc/uaty20/current>

Commercial and free tools intended to help authors to generate accessible PDFs are available. Examples are CommonLook, AxesPDF, or PAVE [12]. PDFAxes has been publicized in several conferences [13-15]. These tools facilitate the preparation of better MS Word documents for accessibility and can also repair the final PDF.

In this paper we contribute to this state-of-the art with the lessons learned in DSAI 2016, enriching the discussion on accessible conference proceedings in [3], and with a suggested integrated workflow for publishing accessible proceedings in an effective and efficient way. We argue in the results and in the discussion that this workflow is needed to smooth the generation of accessible conference proceedings.

3. Case Study: the DSAI 2016 conference

DSAI (Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion) is a conference series devoted to technical innovations on products and services for people with special needs [4]. Since 2006, eight DSAI international conferences have been held. In 2016, DSAI became an ACM in-cooperation conference and its proceedings were published within the ACM Digital Library.

In this paper, we focus on the 7th edition (2016) of the DSAI conference, in which the first author was the publication chair. This edition attracted more than 50 relevant submissions (see table 3).

Table 3 DSAI proceedings number of originals, tables, images and formulas

Main track	Special tracks	Latex originals	MS Word originals	Total pages	Total tables	Total images	Total formulas
16 accepted	43 accepted	11	47	421	97	345	72

3.1 Chosen formats and tools

The ACM submission process defines the format of original authors' submissions and final camera-ready versions of accepted papers. ACM promotes the use of MS Word and LaTeX. ACM provides authors with templates in these formats. ACM sets PDF as a compulsory output format. ACM accepts other output formats (such as XML) as long as a PDF version of the publication is provided too.

The publication chairs of DSAI 2016 wanted to include EPUB format as an additional, more accessible format. Yet, and due to time constraints derived from the publication of the proceedings in an accessible format, the EPUB format was not included. Based on accessibility requirements and the problems documented in the conversion from TEX to PDF, the authors of this paper selected MS Word Office Open XML (.docx) to be the accepted format for original submissions. Thus, Microsoft Office Word 2016 and LaTeX MikTEX were the main authoring tools.

During the conversion procedure, the publication chairs used Adobe Acrobat and Office Axes PDF. The authors also explored two additional tools: Adobe® InDesign, and Infty Editor.

For images, the authors decided to prioritize the JPEG format, as it is the most similar to PDF image internal format and the most adopted one. To edit them Adobe® Photoshop (CC) and GIMP 2.0 were used.

The publication chairs reviewed the accessibility of the papers submitted to the conference by using the MS Word wizard, Adobe® Acrobat DC accessibility validation tool, and PDF Accessibility Checker 2.0.

3.2 Examples and thoughts during the conversion process

Fifty-nine (59) papers were submitted to DSAI. These papers served as a testbed for the conversion tools and the conversion procedure. In this section, some particular problems and the output of the process are discussed.

First of all, the authors experienced difficulties which were unrelated to accessibility. Papers were submitted with an incorrect number of pages. Mandatory elements, such as DOI number, were missing too. Some authors asked us to make some changes in the text after the review process. The version control was not fluid. The authors end up working on old versions of the papers, especially in Special

Tracks. This was a big issue, as it meant going forward and backward from the conversion process and carrying out over and over the same steps.

The authors processed the pictures provided in the submissions by using Gimp or Photoshop to deliver the appropriate resolution or to create a unique image from several grouped images. Three-hundred-and-forty-five (345) images were given alternative texts. Although the authors of the original submissions were informed to deliver alternative texts, the overwhelming majority of them did not provide us with this material, especially when the deadlines for submission were approaching. Consequently, initial conversion trials were very time consuming.

Some single pictures were in fact a collage of pictures. In this case, the resolution of the picture was increased and the collage pictures were joined into one.



Fig 1. Original, composed by two images.

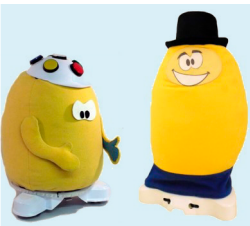


Fig 2 Final, consisting in only one image. Source [21]



Fig 3. Original, composed by 4x6 images.



Fig 4. Final, consisting in only one image. Source [22]

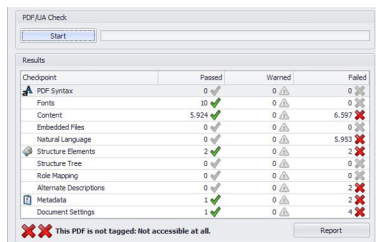
Seventy-two (72) formulas were rewritten with MS Word formula editor, which is not particularly usable because it does not offer a straightforward editor. The symbols have to be added to the formula one by one.

$$\theta_t = \text{atan2} \left(\sqrt{Z^2 + Y^2}, X \right)$$

Fig 5. Example formula

Once converted, the formulas were manually tagged as such in the final PDF.

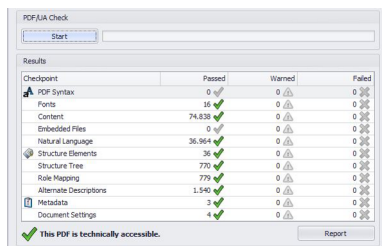
Additionally, a manual review of the final PDF was done to check structure and reading order.



Checkpoint	Passed	Warned	Failed
PDF Syntax	0	0	0
Fonts	10	0	0
Content	5,924	0	6,597
Embedded Files	0	0	0
Natural Language	0	0	5,993
Structure Elements	2	0	2
Structure Tree	0	0	0
Role Mapping	0	0	0
Alternate Descriptions	0	0	2
Metadata	1	0	2
Document Settings	1	0	4

❌ This PDF is not tagged: Not accessible at all.

Fig 6. Example of PDF validation of author's generated document with PDF Accessibility Checker. Several requirements are not meet and do not validate.



Checkpoint	Passed	Warned	Failed
PDF Syntax	0	0	0
Fonts	16	0	0
Content	74,838	0	0
Embedded Files	0	0	0
Natural Language	36,984	0	0
Structure Elements	50	0	0
Structure Tree	770	0	0
Role Mapping	779	0	0
Alternate Descriptions	1,940	0	0
Metadata	1	0	0
Document Settings	4	0	0

✅ This PDF is technically accessible.

Fig 7. PDF validation of the same paper after the publishing process was finished with PDF Accessibility checker. All the requirements are met and validate.

As already mentioned, 2016 was the first year the DSAI proceedings were published in the ACM DL. The committee did not know exactly the procedure to create, for instance, the subjects' metadata and to include copyright information. As a result of this, the required text box with the copyright information did not fit in the initial submissions. Thus, the publication chair had to change the layout / template of the submissions – and conduct the generation and validation of accessibility again.

Eleven (11) papers were converted from LaTeX to MS Word. Fifty-nine (59) papers were reviewed in MS Word by normalizing reading order, author and title information, affiliation format, recreation of subject matter in the correct format, heading hierarchy, table structure and alternative texts. Adobe PDF Accessibility tools and Access PDF validations follow up.

Fifty-nine (59) accessible papers were created and published in an accessible format in the ACM Digital Library. These papers were also published in a complete proceeding book. Although the process was not perfect, and there may be some problems in the reading of formulas and with some alternative text of the pictures, compared to previous editions, DSAI 2016 reached a high level of accessibility in its proceedings, with all papers of the conference being PDF/UA compliant.

4. Lessons learned and a proposal for publishing accessible proceedings

This experience allowed us to understand that the traditional submission process is not suitable for creating accessible proceedings. Achieving this goal rests upon editorial work, additional steps and extra time, and a set of tools.

There should be specific guidelines for authors, additional requirements for images, formulas and tables, and better templates. Most authors have learnt to use the tools on a WYSWYG (what you see is what you get) basis and they rely on the final appearance. Most of them are not aware of accessibility requirements and how to meet them. Accessibility information has not clear affordances, and for example, it is not easy to see if an image has an alternative text or not. Accessible table requirements which imply simple and very structured tables are against widespread authoring practices.

There should be clear steps in the submission process that take into account the accessible reformulation, and reduce changes during the process. With current tools, as we have seen, the conversion and validation procedures require a lot of time. For this process to be efficient, it is necessary to start from a final and correct document.

A new figure in the conference organization is, drawing upon our experience, needed. This figure validates English correctness, length of the papers, required elements for accessibility, layout, and template guidelines before starting the publishing process. Failures to consider these aspects may require posterior editing of the submission and if the accessibility conversion has started it may imply repeating conversion and validation steps.

The authors should have the option to review their paper before the final camera-ready version, because sometimes the accessibility conversion may change a bit the layout or the presentation of images or tables, for example.

In an attempt to contribute to generate more accessible conference proceedings, the authors suggest: (i) creating a new role in the conference organization that takes on the role of accessibility and publishing, (ii) reviewing templates sent to authors and information required from them; and (iii) a clear communication strategy with authors.

4.1 New role and new submission process

The authors suggest that a publication and accessibility chair could be the person in charge of the publication process and of the final editing of the proceedings. This role will focus on all the steps related to the accessible publication of the conference proceedings, ranging from writing the instructions for authors and reviewing submissions to converting the submissions into an accessible version and conducting the accessibility validation.

The authors also suggest adding a new step to the submission process, just after the scientific acceptance of a paper. This new step is a formal review of the accessibility of the paper and, perhaps, should be a requirement for the final decision (accept or reject). After the accessibility requirements are fulfilled, the original paper is processed by the publication chair in order to generate the final paper (see Figure 8).

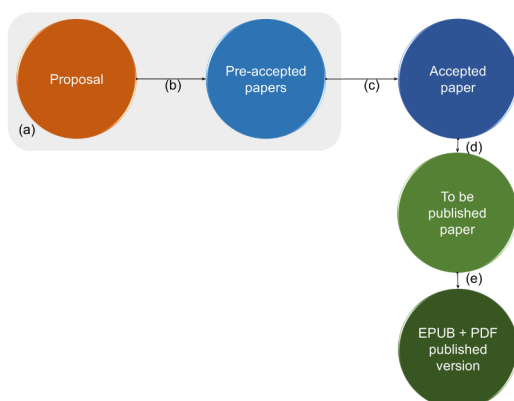


Figure 8. A proposed submission process, the classical procedure (a) of submission and scientific review (b), is enriched with (c) Editorial review (d) Accessible reformulation and (e) author's validation to obtain an EPUB and PDF published version.

4.2 Review of templates and information required to authors

The template for the conference should be reviewed in accessibility terms. The template should help the authors to create accessible documents. Working towards this end, the following information could be provided by the authors:

- the original document in MS Word or Latex
- a document with a brief abstract (2 lines) including the most relevant data from any table included in the paper. This document is needed to create accessible tables.
- a document with the Latex code of any formula included in the paper
- a document with a very short description (125 characters) for each picture or diagram included in the paper. This document should be provided in order to review its content and that it is different from the caption of the picture or diagram.

It is also important to provide a tutorial on how to create more accessible originals, since authors need support. This tutorial (a document or a video) will explain the authors how to configure the correct settings in MS Word and how to add alternative texts. This tutorial will give instructions for image resolution and simplification of tables and on how to write link anchors, describing best practices related to plain language and optimal images.

4.3 Effective communication with authors

The template and the new role in the organizing committee are not enough. From our experience, it is very important to establish a fluent and effective communication with the authors of the papers, because some changes are made in their manuscripts during the process and they should review and confirm them, and also to collaborate in this process.

Before the submission, all authors should be informed about the intention of the conference committee to deliver accessible publications and about some additional requirements for the submissions. A template of the letter to the authors is provided:

Dear Conference author,

this year the Conference Organization Committee is working on the accessibility of the proceedings, and therefore we ask you to perform additional tasks to achieve this important objective.

First of all, you must prepare your images in separate files, named image1.jpg, image2.jpg... with a minimum resolution of 300 dpi.

Moreover, we ask you to try to avoid footnotes. We also ask you to submit 3 additional documents with important information:

- *if your paper includes tables, you must deliver a document tables.pdf with a brief abstract (2 lines) including the most relevant data from the table.*
- *If your paper includes formulas, you must number them in the article, and also you must deliver a document formulas.pdf with the Latex code of the formula*
- *Finally, if you include images, you must deliver an additional document image.pdf, with a very short description (120 characters) for each of them.*

Thanks for your collaboration. Looking forward to meeting you at The Conference.

Best Regards,

XX XX (Conference Chair)

XX XX (Conference Publication Chair)

After the initial publishing process, some changes to the original version of the papers can be made. Thus, authors should be informed about the changes and provide them with the possibility to review them:

Dear Conference author,

We have reached the final phase before the publication of the accepted papers.

All papers have been submitted to a semi-automated process to make them accessible.

The process has introduced the following changes:

- *We have arranged authors reading order to fit visual order*
- *We have deleted false template information in metadata*
- *We have included title, author, subject in metadata*
- *We have included author's alternative text to figures and if we didn't have one we put "See caption" as alternative text*
- *We have reviewed the layout*
- *We have reviewed the heading structure*
- *In some cases we have done minor typo corrections*
- *We have changed language to American English, standard for ACM*

Please verify the final version of your paper(s) available at:

[URL]

If you have any comment on the final version(s) please reply to this email before XXX.

A final reminder for the CCS and alternated text of your paper. This information is very important for the paper accessibility. If you haven't sent this information, please send it in reply to this message to be included in the final version of your paper.

Thanks for your collaboration.

Looking forward to meeting you at The Conference.

Best Regards,

XX XX (Conference Chair)

XX XX (Conference Publication Chair)

A final e-mail is also necessary to keep authors abreast of aspects which could not be dealt with in the current submission and that could be considered for future manuscripts (see letter X in Annex I):

Dear author,

After reviewing the accessibility of your paper, we found (and repaired) the following issues:

[] some images lack alternative text

[] some images have a poor color contrast

[] some tables lack alternative text

[] some links did not have a meaningful anchor text

[] some tables were very complex and not easily linearized

In future editions we will appreciate if you fulfill these requirements in the original submission.

Yours truly,

Document chair

5. Discussion

The first question that arises from our experience is whether it is sensible to make such a big effort to publish accessible proceedings. The authors' answer is yes. Accessibility is a right and conference proceedings are an important source of scientific progress. Nevertheless, accessible publishing is not possible without a collective effort involving all the stakeholders, companies that create authoring tools, big publishers, conference organizers, authors... A big claim should be made by scientific community to have all the stakeholders engaged in accessibility. As we have seen, there is a correct and updated set of standards for accessible PDF (PDF/UA) and for EPUB (EPUB3) which include accessibility requirements. Unfortunately, authoring tools and widespread authoring practices have not kept pace with this evolution, and this is needed, as we have shown.

There is also room for thinking whether this case study has chosen the best tools for the process and if it is generalizable to other conferences. In this case, the answer is a partial yes, because it has focused only on textual documents, which are the main publishing format in conferences. The case study did not involve submissions including video, data or code content, which are more and more common nowadays. This heterogeneity can be seen as another opportunity to work on and improve the accessibility of conference proceedings.

6. Conclusion

The main conclusion of our experience is that with the current authoring tools, it is not realistic – or even fair – to rely solely on authors to create truly accessible PDF papers. Researchers need better tools to write and generate accessible PDF documents.

PDF authoring and conversion tools are not mature enough for an easy publication workflow. They require expert validation and review of converted documents. This article described non-evident limitations of the conversion process, and highlighted the semantic reduction of final tags. The final PDF is not yet optimized for viewing in tablets or mobiles.

Although there is also a lack of EPUB3 authoring tools, the transparency of this format may facilitate a better authoring and it may be desirable to deliver EPUB3 files in addition to the required PDF.

There is a need for an integrated workflow of generating accessible conference proceedings, which, the authors suggest, consists of four key elements: better tools, better communication between conference organization and researchers, better templates and instructions, and more awareness of accessibility.

7. Future research

The conversion from Latex and MSWord to PDF and EPUB needs further research. SIGCHI and TUG may offer some solutions.

Apart from basic textual proceedings, making video, data and code more accessible in conference proceedings should be taken into consideration in near future.

More research is needed to test resulting proceedings with assistive tools and people with disabilities to prioritize accessibility efforts.

The authors are currently working in improving Pandoc conversion from TEX to HTML, in order to be able to produce EPUB3 output from TEX sources.

References

- [1] J. Lazar *et al.*, Making the field of computing more inclusive, *Commun. ACM*, vol. 60, no. 3, pp. 50–59, Feb. 2017.
- [2] J. Mankoff, The wicked problem of making SIGCHI accessible, *Interactions*, vol. 23, no. 3, pp. 6–7, 2016.
- [3] E. Brady, Y. Zhong, and J. P. Bigham, Creating accessible PDFs for conference proceedings, in *Proceedings of the 12th Web for All Conference on - W4A '15*, 2015, pp. 1–4.
- [4] TUG, Accessibility and PDF Standards - TeX Users Group. [Online]. Available: <http://www.tug.org/twg/accessibility/>. [Accessed: 27-Feb-2017].
- [5] P. Krautzberger, What's the best TeX-to-HTML or TeX-to-ePUB converter? · Peter Krautzberger, 2013. [Online]. Available: <https://www.peterkrautzberger.org/0127/>. [Accessed: 03-Jan-2017].
- [6] É. P. F. de L. (EPFL), [DH101] TeX to Ebook Converter | A tool to convert LaTeX documents to ebook formats (EPUB, Mobipocket)., 2013. [Online]. Available: <https://tex2ebook.wordpress.com/>. [Accessed: 03-Jan-2017].
- [7] T. E. D. M. L. (EuDML), Tools & Technical Specifications | EuDML [Online]. Available: <https://initiative.eudml.org/tools-technical-specifications>. [Accessed: 03-Jan-2017].
- [8] N. I. of S. and T. (NIST), LaTeXML A LaTeX to XML/HTML/MathML Converter, 2016. [Online]. Available: <http://dlmf.nist.gov/LaTeXML/>. [Accessed: 03-Jan-2017].
- [9] S. Nesbitt, Writing and publishing an ebook with LaTeX, pandoc, and calibre · Open Source Musings, 2015. [Online]. Available: <http://scottnesbitt.io/2015/10/05/ebook/>. [Accessed: 03-Jan-2017].
- [10] J. MacFarlane, Pandoc - Creating an ebook with pandoc, 2016. [Online]. Available: <http://pandoc.org/epub.html>. [Accessed: 03-Jan-2017].
- [11] J. P. Bigham, E. L. Brady, C. Gleason, A. Guo, and D. A. Shamma, An Uninteresting Tour Through Why Our Research Papers Aren't Accessible, *Proc. 2016 CHI Conf. Ext. Abstr. Hum. Factors Comput. Syst. - CHI EA '16*, pp. 621–631, 2016.
- [12] K. McCall and K. Communications, Accessible PDF Remediators Survey Results 2017”
- [13] R. Bianchetti, M. Erle, and S. Hofer, Mainstreaming the creation of accessible PDF documents by a rule-based transformation from word to PDF, in *ICCHP. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7382 LNCS, no. PART 1, pp. 595–601.
- [14] R. Bianchetti, S. Hofer, and M. Erle, Correcting ‘ Last Mile ’ Errors - Quality Assurance of PDF / UA Documents without Being a Developer Introduction : No Authoring Software Yet with Full Support for PDF / UA The Concept : Defining ‘ Last Mile ’ Errors and Developing, in *ICCHP*, 2014, pp. 193–196.
- [15] R. Bianchetti, S. Hofer, and M. Erle, Implementing PDF / UA in Microsoft Word - How Can PDF / UA Become an Everyday Part of Document Authoring ? Introduction : Is PDF / UA Only for Insiders , Geeks and Experts ? The Concept : A New Role Based Model - Only the Template Creator Has to Be the PDF in *ICCHP*, 2014, pp. 181–184.
- [16] T. Raman, Audio System for Technical Readings. Dissertation. Cornell University, 1994.
- [17] M. Ribera, Are PDF Documents Accessible? *Information Technology & Libraries* 27.3 (2008): 25–43. Web
- [18] E. systems and authenticity of information Technical Committee : ISO/TC 171/SC 2 Document file formats, ISO 32000-2017: Document management - Portable document format - Part 2: PDF 2.0, 2017.
- [19] D. Gunn, Accessible eBook Guidelines for Self-Publishing Authors Foreword, 2016.
- [20] J. Roig and M. Ribera, Creation of accessible EPUB documents by non-technical users: A long way to go, in *ACM International Conference Proceeding Series*, 2015, vol. 07–09–Sept.
- [21] A. Bonarini [et al.] Playful interaction with Teo, a mobile robot for children with neurodevelopmental disorders in *DSAI 2016*, 2016, p.225
- [22] M. Farrajota, J.M:F Rodrigues and J.M.H du Buf Using multi-stage features in fast R-CNN for Pedestrian Detection in *DSAI 2016*, 2016, p.406